Dissertations of Department of Applied Physics, Helsinki University of Technology,
Teknillisen fysiikan laitos, Teknillinen korkeakoulu
Dissertation 154
Espoo 2009

COMPUTATIONAL STUDIES OF III-V COMPOUND SEMICONDUCTORS

Katri Laaksonen

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Faculty of Information and Natural Sciences for public examination and debate in Auditorium K at Helsinki University of Technology (Espoo, Finland) on the 23rd of January, 2009, at 13 o´clock.

Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Applied Physics

Teknillinen korkeakoulu
Informaatio- ja luonnontieteiden tiedekunta
Teknillisen fysiikan laitos

Dissertations of Department of Applied Physics
Helsinki University of Technology
ISSN 1797-9595 (print)
ISSN 1797-9609 (online)

Dissertation 154 (2009):
Katri Laaksonen: Computational studies of III-V compound semiconductors

Opponent:
Dr. Susanne Mirbt, Uppsala University, Sweden

Pre-examiners:
Docent Kalevi Kokko, University of Turku, Finland
Dr. Christopher Castleton, Nottingham Trent University, United Kingdom

ISBN 978-951-22-9723-8 (print)
ISBN 978-951-22-9724-5 (online)

Picaset Oy
Helsinki 2009

ABSTRACT OF DOCTORAL DISSERTATION	HELSINKI UNIVERSITY OF TECHNOLOGY P. O. BOX 1000, FI-02015 TKK http://www.tkk.fi

Author Katri Laaksonen

Name of the dissertation

Computational studies of III-V compound semiconductors

Manuscript submitted 14.10.2008	Manuscript revised 16.12.2008
Date of the defence 23.1.2009 at 13 o'clock, TKK main building, Otakaari 1, lecture hall K	
☐ Monograph	☒ Article dissertation (summary + original articles)

Faculty	Faculty of Information and Natural Sciences
Department	Department of Applied Physics
Field of research	Computational materials physics
Opponent(s)	Dr. Susanne Mirbt, Uppsala University
Supervisor	Acad. Prof. Risto Nieminen
Instructor	

Abstract

Nitrogen-containing III-V compound semiconductors are uniquely suited for many applications in electronics and optoelectronics. For example, the wide-band-gap semiconductor GaN and its alloys are essential for many recently developed optoelectronic devices such as white light emitting diodes (LEDs) and lasers with colors from red to ultraviolet.

The electrical and optical properties of semiconductor materials are mainly determined by a small number of defects and impurities that have been intentionally or unintentionally incorporated into the material. In a compound semiconductor the compositional uniformity can also have an important effect on the electronic properties of the alloy. In this thesis these issues are studied in III-V compound semiconductors with computational modeling.

A density-functional theory (DFT) based method is used to determine the structure and energetics of 1) vacancies and substitutional As and In defects in GaN, 2) N interstitial defects in GaAs, 3) vacancies in AlN and 4) GaAsN and GaInN alloys. The results from the defect calculations for the various systems include the most prevalent types of the defects and their most important properties under different growth conditions.

Compositional instability in GaInN is studied with multiscale modeling where interaction parameters for the lattice kinetic Monte Carlo simulations of the alloy decomposition are taken from the DFT calculations. The simulations show the high sensitivity of the GaInN decomposition to relatively small variations of the interaction between Ga and In atoms and offer an explanation for the diversity of the alloy decomposition patterns observed in the experiments.

Keywords	III-V semiconductors, defects, density functional theory		
ISBN (printed)	978-951-22-9723-8	ISSN (printed)	1797-9595
ISBN (pdf)	978-951-22-9724-5	ISSN (pdf)	1797-9609
Language	English	Number of pages	47 p. + app. 59 p.
Publisher	Department of Applied Physics, Helsinki University of Technology		
Print distribution			
☒ The dissertation can be read at http://lib.tkk.fi/Diss/2009/isbn9789512297245/			

VÄITÖSKIRJAN TIIVISTELMÄ	TEKNILLINEN KORKEAKOULU PL 1000, 02015 TKK http://www.tkk.fi

Tekijä Katri Laaksonen

Väitöskirjan nimi

III-V-yhdistepuolijohteiden laskennallisia tutkimuksia

Käsikirjoituksen päivämäärä 14.10.2008	Korjatun käsikirjoituksen päivämäärä 16.12.2008
Väitöstilaisuuden ajankohta 23.1.2008 klo 13, Teknillisen korkeakoulun päärakennus, Otakaari 1, sali K	
☐ Monografia	☒ Yhdistelmäväitöskirja (yhteenveto + erillisartikkelit)

Tiedekunta	Informaatio- ja luonnontieteiden tiedekunta
Laitos	Teknillisen fysiikan laitos
Tutkimusala	Laskennallinen materiaalifysiikka
Vastaväittäjä(t)	Dr. Susanne Mirbt, Uppsala University
Työn valvoja	Akatemiaprofessori Risto Nieminen
Työn ohjaaja	

Tiivistelmä

Typpeä sisältävät III-V-yhdistepuolijohteet soveltuvat erittäin hyvin moniin elektroniikan ja optoelektroniikan sovelluksiin. Esimerkiksi leveän energia-aukon puolijohde GaN ja sen seokset ovat välttämättömiä monissa äskettäin kehitetyissä optoelektronisissa laitteissa kuten valkoisissa valodiodeissa (light emitting diode, LED) sekä lasereissa, joiden väri on punaisen ja ultavioletin väliltä.

Puolijohdemateriaalien sähköiset ja optiset ominaisuudet määrää pääosin pieni määrä kidevirheitä ja epäpuhtauksia, jotka ovat tarkoituksella tai tahattomasti sekoittuneet materiaaliin. Yhdistepuolijohteissa myös koostumuksen tasaisuudella voi olla suuri vaikutus seoksen elektronisiin ominaisuuksiin. Tässä väitöskirjassa näitä kysymyksiä tutkitaan III-V-yhdistepuolijohteissa laskennallista mallinnusta käyttäen.

Tiheysfunktionaaliteoriaan (density functional theory, DFT) perustuvaa menetelmää käytetään 1) GaN:n vakanssien sekä As- ja In-korvaushilavirheiden, 2) GaAs:n typpivälisijojen, 3) AlN:n vakanssien sekä 4) GaAsN- ja GaInN-seosten rakenteiden ja energioiden määrittämiseen. Eri systeemien hilavirhelaskujen tuloksiin sisältyvät yleisimmät hilavirhetyypit sekä niiden tärkeimmät ominaisuudet eri kasvatusoloissa.

Koostumuksen tasaisuutta GaInN:ssä tutkitaan monimittakaavamallinnuksella, missä hilakineettisellä Monte Carlo (lattice kinetic Monte Carlo) -menetelmällä tehdyissä seoksen hajaantumisen simulaatioissa käytetyt vuorovaikutusparametrit saadaan DFT-laskuista. Simulaatiot osoittavat, kuinka herkkä GaInN:n hajaantumistapa on verrattain pienille muutoksille Ga- ja In-atomien vuorovaikutuksessa, mikä selittää kokeissa havaitun hajaantumistapojen vaihtelevuuden.

Asiasanat III-V-puolijohteet, hilavirheet, tiheysfunktionaaliteoria

ISBN (painettu) 978-951-22-9723-8	ISSN (painettu) 1797-9595
ISBN (pdf) 978-951-22-9724-5	ISSN (pdf) 1797-9609
Kieli englanti	Sivumäärä 47 s. + liitt. 59 s.

Julkaisija Teknillisen fysiikan laitos, Teknillinen korkeakoulu

Painetun väitöskirjan jakelu

☒ Luettavissa verkossa osoitteessa http://lib.tkk.fi/Diss/2009/isbn9789512297245/

Preface

This thesis has been prepared in the Computational Nanoscience group (COMP) of Laboratory of Physics, Helsinki University of Technology (TKK), during years 2003-2008.

I wish to express my gratitude to my supervisor Academy Professor Risto Nieminen as well as Dr. Maria Ganchenkova (TKK), Dr. Hannu-Pekka Komsa and Professor Tapio Rantala (Tampere University of Technology) with whom I have mainly worked with during the preparation of this thesis. I am also grateful to many other members of the staff of Laboratory of Physics who have contributed to my work through many discussions over the years. I would especially like to thank Dr. Paula Havu, Dr. Karen Johnston and M.Sc. Laura Oikkonen with whom I have had the pleasure to share an office as well as thoughts on many subjects, including (surprisingly often) even science.

The financial support from the Emil Aaltonen foundation and the computational resources provided by CSC, the Finnish IT center for science, are gratefully acknowledged.

Finally I would like to thank my family and Timo for all their support and encouragement.

Espoo, September 2008

Katri Laaksonen

Contents

List of publications

This thesis consists of an overview and the following publications:

I K. Laaksonen, H.-P. Komsa, E. Arola, T. T. Rantala and R. M. Nieminen, *Computational study of $GaAs_{1-x}N_x$ and $GaN_{1-y}As_y$ alloys and arsenic impurities in GaN*, J. Phys.: Condens. Matter **18**, 10097 (2006)

II K. Laaksonen, M. G. Ganchenkova and R. M. Nieminen, *Minor component ordering in wurtzite $Ga_{1-x}In_xN$ and $Ga_{1-x}Al_xN$*, Physica B **376-377**, 502 (2006)

III M. G. Ganchenkova, V. A. Borodin, K. Laaksonen and R. M. Nieminen, *Modeling the compositional instability in wurtzite $Ga_{1-x}In_xN$*, Phys. Rev. B **77**, 075207 (2008)

IV K. Laaksonen, H.-P. Komsa, T. T. Rantala and R. M. Nieminen, *Nitrogen interstitial defects in GaAs*, J. Phys.: Condens. Matter **20**, 235231 (2008)

V K. Laaksonen, M. G. Ganchenkova and R. M. Nieminen, *Vacancies in wurtzite GaN and AlN*, J. Phys.: Condens. Matter **21**, 015803 (2009)

The author has had an active role in all the phases of the research reported in this thesis. She has been involved in planning and performing the ab-initio calculations as well as in the analysis of the results in Publication I (especially the N-rich GaAsN alloys), Publication II, Publication III (ab-initio results), Publication IV and Publication V (AlN). She has written the main drafts of Publications I, II, and IV and contributed to the writing of Publications III and V.

1 Introduction

Semiconductor materials are essential for modern electronics. Many semiconductor based devices (such as computers and mobile phones) have become an inseparable part of everyday life. The best known and most widely used semiconductor material is silicon but many others have important applications as well. Wide-band-gap semiconductors are needed for light emitting devices operating in the blue or ultraviolet spectral regions, such as white light emitting diodes (LEDs) and blue lasers (for example, for high-density optical data storage in discs similar to DVDs) which can already be found in many homes. One such material is the wide-band-gap semiconductor gallium nitride (GaN) which together with its alloys and other III-V compound semiconductors has been studied in this thesis.

The electrical conductivity of a semiconductor material can be controlled by the introduction of small concentrations of impurities in the material in a process called doping. For the production of electronic devices, the ability to dope the material both n-type, where the electrical carriers are electrons, and p-type, where they are holes, is necessary but can be complicated or prevented by impurities or native defects such as vacancies (atom missing from a lattice site) or interstitials (atom on an off-lattice site). These point defects can act as compensating centers that deactivate intentionally incorporated donor or acceptor impurities. They can also be optically active and affect the optical properties of the materials. Therefore, it is of high importance to gain knowledge on the formation of defects.

In this thesis defects in III-V compound semiconductors are studied using electronic-structure calculations based on density functional theory (DFT). The method is well-suited for studying the effects of defects on the electronic and structural properties of semiconductor materials. With the ongoing development of algorithms, methods and computers DFT calculations can be used to explain, interpret and even predict the properties of materials and phenomena observed in experiments.

The defects considered in this thesis include arsenic (As) defects in GaN (Publication I) where the aim is to clarify the experimental and theoretical results that try to determine which substitutional site (Ga or N) is preferred by As impurities in GaN. In Publication IV the opposite end of the alloy composition range is considered. Nitrogen is known from previous studies to prefer the isoelectronic substitutional As-site in GaAs. However, interstitial

N has also been observed in experiments and it too can have an important effect on the electronic properties of the dilute nitrogen GaAsN alloys. Publication V deals with vacancies in aluminium nitride (AlN) and GaN which, despite being very simple defects in widely studied materials, still offer open questions for studies as well as challenges for the methodology used in the determination of the formation energies and transition levels.

The computational method used in these studies has some limitations, some of which arise from the DFT method itself and others from the limited computing resources. For example, in the above mentioned studies of defects in bulk semiconductors, the aim is to simulate an isolated defect in a pure host as accurately as possible despite the periodic boundary conditions (PBC) which create an infinite array of interacting defect images. The obvious solution would be to use supercells large enough for even long-range interactions to become negligible but unfortunately the current methods and resources allow calculations with supercells containing at most hundreds of atoms. These limitations of the computational method are discussed in this thesis together with possible solutions. Similarly the attainable supercell sizes restrict the simulation of truly random alloys or ordering processes since the periodic images create artificial order in the system.

The III-V compound semiconductors are often alloyed with other group-III or -V elements in order to produce materials with desired properties such as energy gaps suitable for producing light of a required wavelength. One of the alloys considered in this thesis is GaAsN (Publication I) where the concentrations of the minority component (As or N) are restricted to a few per cent. In this case, the consequences of the alloying are deduced from calculations of the various configurations of a small number of impurity atoms in a relatively small supercell (essentially ordered alloys). However, for the simulation of gallium indium nitride (GaInN) alloys with far higher concentrations of In, a different approach is used which goes beyond DFT calculations.

2 Methods

2.1 Density functional theory

Density functional theory (DFT) is one of the most successful methods for performing electronic-structure calculations in condensed matter physics and other fields. It is a "nearly" ab-initio (first principles) method which means that it does not contain any parameters that have been adjusted to obtain agreement with experimental results. It offers a good balance between accuracy and computational efficiency which has led to its popularity. The theory behind DFT and the methods used in practical calculations are presented in this chapter while the next one deals with the application of the method to the studies of defects in semiconductors with high accuracy.

2.1.1 Hohenberg-Kohn theorem

The total many-particle wavefunction of a system with N electrons depends on $3N$ variables. This would make computing the properties of the system from the many-particle Schrödinger equation too heavy for most systems with more than a few electrons. The main idea behind DFT comes from the theorem by Hohenberg and Kohn [1]. The first statement of the theorem states that all ground-state properties of a many-particle system are uniquely determined by the ground-state electron density $n(\mathbf{r})$. The tremendous simplification that follows from being able to deal with the density instead of a wavefunction with $3N$ variables makes it possible to study, among others, bulk systems with a high number of electrons. According to the second statement of the theorem there exists a universal functional for the energy $E[n(\mathbf{r})]$ in terms of the density, valid for any external potential $V_{\text{ext}}(\mathbf{r})$. For a given $V_{\text{ext}}(\mathbf{r})$, the density that minimizes the functional is the ground-state density.

2.1.2 Kohn-Sham equations

The Kohn-Sham equations allow DFT to be used in practical calculations for real-life systems [2, 3]. They transform the problem of interacting electrons in an external ion potential to the problem of non-interacting electrons

moving in an effective potential. The total energy of the system can then be calculated from

$$E[n(\mathbf{r})] = E_{\text{kin}}[n(\mathbf{r})] + E_{\text{H}}[n(\mathbf{r})] + \int n(\mathbf{r})V_{\text{ext}}(\mathbf{r})\mathrm{d}\mathbf{r} + E_{\text{xc}}[n(\mathbf{r})]. \quad (1)$$

$E_{\text{kin}}[n(\mathbf{r})]$ is the kinetic energy and $E_{\text{H}}[n(\mathbf{r})]$ the Coulomb energy (also called the Hartree energy) of the non-interacting electrons. The first three terms can all be dealt with accurately and relatively easily. The last term, which contains the remaining energy contributions, is called the exchange-correlation energy. The exchange energy follows from the Pauli exclusion principle which prevents two identical fermions from occupying the same quantum state simultaneously. The correlation energy comes from the electron-electron Coulomb repulsion which leads to a reduction in probability of an electron being near another electron. If the exchange-correlation functional was known, all of the system properties could be calculated exactly. In practice, however, approximations need to be used and they are discussed further in section 2.1.3.

Self-consistent iteration can be used to solve the Kohn-Sham equations

$$[-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r})]\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r}) \quad (2)$$

where the density is expressed in terms of N occupied single-particle orbitals $\phi_i(\mathbf{r})$ (with i denoting the states)

$$n(\mathbf{r}) = \sum_i^N |\phi_i(\mathbf{r})|^2 \quad (3)$$

and the effective potential is

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{\mathbf{r} - \mathbf{r}'}\mathrm{d}\mathbf{r}' + \frac{\partial E_{\text{xc}}[n(\mathbf{r})]}{\partial n(\mathbf{r})}. \quad (4)$$

2.1.3 Exchange and correlation functionals

Density functional theory would, in principle, be exact if the exact form of the exchange-correlation functional (the "divine" functional) were known.

Unfortunately, this is not the case, at least currently, so in practice approximations need to be relied on, some of which are shortly presented in the following.

Local (spin) density approximation (LDA or LSDA for the spinpolarized form) relates the exchange-correlation energy of a non-uniform system locally to that of a uniform electron gas with the same density $n(\mathbf{r})$:

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[n(\mathbf{r})] = \int n(\mathbf{r})\epsilon_{\mathrm{xc}}^{\mathrm{LDA}}(n(\mathbf{r}))\mathrm{d}\mathbf{r} \tag{5}$$

where $\epsilon_{\mathrm{xc}}^{\mathrm{LDA}}(n(\mathbf{r}))$ is the exchange-correlation energy density of the homogeneous electron gas. LDA tends to overbind molecules and solids and to underestimate bond and lattice distances [4] but, considering its simplicity, it works surprisingly well for many systems and is still widely used in electronic-structure calculations in solid state physics [5]. However, if one wishes to take the next step towards the exact exchange-correlation functional, one seemingly natural way of doing it is to include the gradients of the density which leads to the so-called generalized gradient approximation (GGA):

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[n(\mathbf{r})] = \int f(n(\mathbf{r}), \nabla n(\mathbf{r}))\mathrm{d}\mathbf{r}. \tag{6}$$

Several different forms of GGA have been proposed so far with some of the most common being formulated by Perdew and Wang (PW91) [6] and Perdew, Burke and Ernzerhof (PBE) [7, 8]. Compared to LDA, GGA generally (but not always, especially for solids) improves (and sometimes overcorrects) the ground state energies, bond lengths and dissociation energies of molecules and cohesive properties of solids [6].

One of the failures of the Kohn-Sham DFT is the well-known band-gap problem which also affects the defect studies in semiconductors. With standard DFT the difference of the energies of the highest occupied and lowest unoccupied Kohn-Sham orbitals (called the Kohn-Sham band gap in the following) can be significantly smaller than the fundamental band gap E_{g}, defined as the difference between the ionization potential I and the electron affinity A of the bulk material

$$E_{\mathrm{g}} = I - A. \tag{7}$$

In some cases (such as InN [9]) the band gap of a material experimentally known to be a semiconductor or insulator can even vanish in a DFT calculation. In order to gain more understanding of the reasons behind the band-gap problem, two of the more important formal deficiencies of LDA and GGA are considered in the following.

The potential should exhibit a discontinuity: its absolute value should equal the ionization potential I if the integer number of electrons is approached from below, but it should equal the electron affinity A if the number of electrons is approached from above. If the total energy were plotted as a function of particle number (allowing also fractional charges), the curve would be continuous, but its derivative would have kinks at the integer values of the occupation of the orbitals [10, 11]. Considering the different terms of the total energy equation (1), the derivatives of the external potential and the Hartree terms are continuous with respect to density while the non-interacting kinetic energy and the exchange-correlation energy should exhibit discontinuities in their derivatives. Unfortunately, one of the deficiencies of the LDA and GGA potentials is that they do not exhibit particle number discontinuity. However, the difference between the Kohn-Sham band gap and the fundamental band gap has its source in the Kohn-Sham approach itself and even if the exact exchange-correlation functional with the derivative discontinuity was used, the Kohn-Sham gap may differ from the fundamental gap [10, 4].

Another serious deficiency of the LDA and GGA functionals is the presence of self-interaction. In eq. (1) the effects of the electron-electron interaction on the total energy have been divided into a classical Coulomb part and an exchange-correlation part which should contain everything else. This results in the second term of eq. (4) (the Hartree potential), which means that each electron is repelled from the total charge of the system – including spuriously also its own charge. This effect is known as the self-interaction error. Since DFT, in principle, handles the electron-electron interaction exactly, the error made in the Hartree term should be completely canceled out by the exchange-correlation term. However, this is true only for the exact exchange-correlation functional and in LDA and GGA it is only partially canceled. For example, in defect studies problems can arise when LDA or GGA fails quantitatively on localized states, but does better for delocalized states, leading to a distorted overall qualitative picture [10]. Some methods, which go beyond the standard DFT calculations and could be used in defect studies in the future, are discussed in section 3.5.

In order to determine which exchange-correlation functional (LDA or one of the flavors of GGA) is more suitable for the studies presented in this thesis, an extensive comparison of the values given by different approximations and the experiments for the basic properties (such as lattice parameters, cohesive energies and band gaps) was performed. LDA, as parametrized by Perdew and Zunger [12] based on the quantum Monte Carlo calculations by Ceperley and Alder [13], was chosen as the exchange-correlation functional based on the results. For example, in the case of GaAsN alloys it gives a decisively better band gap compared to GGA (PW91), as discussed in Publication I.

2.1.4 Vienna ab-initio simulation package

Based on the DFT theory presented above, a large number of electronic-structure methods have been developed over the last decades to suit a variety of purposes and applications. The calculations presented in this thesis have been performed using the Vienna ab-initio simulation package (VASP) [14] which uses plane waves as the basis functions for the expansion of the Kohn-Sham orbitals.

To describe the rapidly oscillating wavefunctions in the core regions with deep Coulomb potentials, a high number of plane waves is needed. However, since the physical and chemical properties of the atoms depend mainly on the valence electrons, this can be avoided by the use of a weaker pseudopotential which replaces the core electrons and the ionic potential of the nucleus. A pseudopotential is smoother than the true potential in the core region but has the same scattering properties for the wavefunctions in the valence region (see fig. 1) [15].

In some systems, such as many of the III-V compound semiconductors (GaN, GaAs, InN), there are core electrons that are not chemically inert. Instead they are important in chemical bonding and cannot simply be included in the pseudopotential representing the nucleus and core electrons (the frozen core approximation is used). They can be treated via so-called nonlinear core corrections or included explicitly as valence electrons. In the DFT studies in this thesis, these semi-core electrons (Ga $3d$ and In $4d$ electrons) have been treated explicitly. In addition, the projector augmented wave method (PAW) [16, 17] was used. It reconstructs the all-electron wave functions of the electrons treated explicitly in the calculation and offers

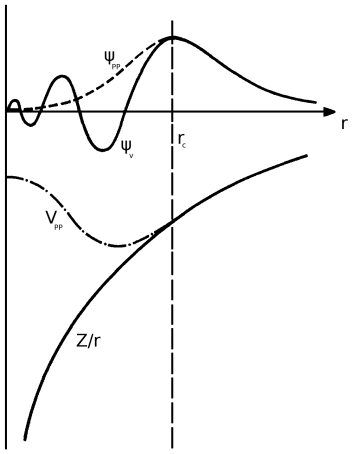

Figure 1: Schematic illustration of the pseudopotential (dash-dot line) and all-electron potential (solid) and the corresponding wavefunctions. Pseudopotential and all-electron potential are equal and give the same wavefunctions outside radius r_c. Based on [15].

a good combination of the efficiency and flexibility of a pseudopotential method with the accuracy of an all-electron method.

In a practical calculation using VASP, the Kohn-Sham equations are first solved numerically through self-consistent iteration starting from some initial trial density. According to the Born-Oppenheimer approximation the electron and ion movements can be decoupled based on the large mass difference. In a defect calculation, the surrounding atoms are affected by the creation of the defect and the ion positions need to be relaxed to reach the minimum total energy. This is done in steps where the ions are moved according to the Hellmann-Feynman forces acting on the ions, with the Kohn-Sham equations solved at each step for the current ion positions.

The forces on the ions are determined using the Hellmann-Feynman theorem because it greatly simplifies the practical calculations. The forces are the negative derivatives of the total energy with respect to ion positions and, in principle, they could be estimated by displacing the ions in all directions and solving the Kohn-Sham equations for each of the configurations. This would, however, require $6N_{\mathrm{I}}$ calculations for N_{I} ions. According to the Hellmann-Feynman theorem [18, 19], if the wavefunctions ψ_i are the eigenstates of

the Hamiltonian H of the system, the electronic degrees of freedom do not contribute to the derivatives of the energy:

$$\frac{dE}{d\mathbf{R}_\mathrm{I}} = \frac{\partial E}{\partial \mathbf{R}_\mathrm{I}} = \frac{\partial < \psi_i | H | \psi_i >}{\partial \mathbf{R}_\mathrm{I}} = < \psi_i | \frac{\partial H}{\partial \mathbf{R}_\mathrm{I}} | \psi_i > . \quad (8)$$

Forces calculated using the Hellmann-Feynman theorem are very sensitive to errors in the wavefunctions, which need to be very close to the exact eigenstates of the Hamiltonian.

Many of the calculational details (such as supercell sizes, cutoff energies for the plane waves and **k**-point samplings for the Brillouin zone integrations) depend on the properties of the system in question and are given for the different systems studied in this thesis in each of the Publications.

2.2 Lattice kinetic Monte Carlo

DFT methods can currently be applied to systems with up to hundreds of atoms but to study effects with longer length-scales, such as the kinetics of the formation of microstructure, a multiscale approach needs to be used. DFT calculations performed with VASP can give reliable formation, binding and migration energies for defects. These results can be used as input parameters to lattice kinetic Monte Carlo (LKMC) calculations. LKMC simulations can be used to study, for example, diffusion, ordering and phase separation for thousands or even millions of atoms in the simulation cell for experimentally relevant annealing times (from microseconds to hours at the temperatures of 300-1000°C). In Publication III this kind of multiscale approach has been used to study the ordering and phase separation in GaInN. The LKMC simulations have been performed with the CASINO-LKMC code [20].

Monte Carlo simulations are stochastic simulations where random numbers are used in the simulation algorithm. In the Monte Carlo methods used in statistical physics (see e.g. [21]) the basic idea is to simulate the random thermal fluctuations of a system which changes from one state to another during an experiment. They work by choosing a subset of states for the system at random from some probability distribution which is specified according to the system under consideration. The Monte Carlo algorithm can be presented in a much simplified form as a sequence of steps which are repeated:

1. One event is chosen from all of the possible events for the system in question.
2. The event is accepted if the (predefined) probability of the event is higher than a generated random number in the range [0,1).
3. The list and probabilities of the possible events are updated based on the possible consequences of the previous step.

The Monte Carlo codes used in practice differ greatly depending on the problem in question and also on the implementation of the algorithm.

The basic events in the simulation can be defined in different ways depending on the system and the process one is interested in. For example, in the case of defect clustering studies in solids, the individual defects and different types of defect complexes can all be considered as different objects. The different objects have different properties and, for example, the properties of the complexes do not need to have anything in common with the properties of the defects from which they are formed. The events in the system can then be defined as tranformations of the objects to different types of objects. This type of simulation can be called object (kinetic) Monte Carlo. Another type of approach to this kind of problem is lattice kinetic Monte Carlo. There the basic events are jumps where two atoms or defects (such as vacancies) exchange positions in a predefined lattice of atom positions (which corresponds to the crystal structure of the material) and the jump probabilities of the atoms or defects can be calculated based on the local environment.

Often the purpose of the simulation is to find the most energetically favorable final state (the equilibrium) and it is not important if the intermediate states are physically unreasonable. However, sometimes the main interest is not in finding the final state (which might never be realized) but the kinetic pathways of, for example, system decomposition. In that case also the event acceptation criteria have to take into account the physics of the individual events at each intermediate step. This kind of simulation technique that follows the physical kinetic pathway is often called "kinetic" Monte Carlo. With the CASINO-LKMC code the kinetic pathways of the system evolution can be studied by taking into account the energetics of the defect migration (energy barriers).

In the CASINO-LKMC code the probabilities of all possible events that can happen at the current step need to be known at the beginning of the step.

On one hand it can slow the simulation down since the probabilities affected by the previous step need to be recalculated before each new step, but on the other hand it can also lead to a significant speed-up if there are many events with very low or vanishing probabilities ("empty" events). In this case the empty events can be discarded already in the beginning of the step rather than repeatedly first being chosen at random and then rejected.

CASINO-LKMC uses the "pair-bond" approximation which assumes that the total energy of the crystal can be presented as a sum of contributions E_{ij} coming from each pair of atoms in the crystal:

$$E = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} E_{ij} \tag{9}$$

where N is the total number of atoms. Usually it is assumed that the contributions E_{ij} depend only on the types of the atoms or defects and the distance between them. Also, since the interaction between the atoms decreases with distance, only atoms up to a certain distance (up to the fourth nearest neighbours in Publication III) are taken into account. Since the distances between the atoms are fixed to certain values in lattice kinetic Monte Carlo, they can be defined as the number of coordination spheres (neighbours) separating the atoms. Then the cutoff distance for the interaction corresponds to the highest possible number of coordination shells between two atoms where the interaction is still taken to be non-zero. The total energy can thus be rewritten as

$$E = \sum_{X} \sum_{Y} \sum_{k=1}^{N} Z_{XY}^{(k)} E_{XY}^{(k)} \tag{10}$$

where $Z_{XY}^{(k)}$ and $E_{XY}^{(k)}$ are the total number and the interaction energy, respectively, for the pairs of atoms of types X and Y located at distance k while the summation is over the types and separations of the atoms. The interaction energies corresponding to the bonds between atoms must be determined from a source that is external to the LKMC simulation, such as experiments or DFT calculations (as in Publication III).

In the beginning of a LKMC simulation there is some initial configuration of impurity atoms and point defects on a small section of the lattice of the material (the simulation cell). The jump frequencies are then calculated

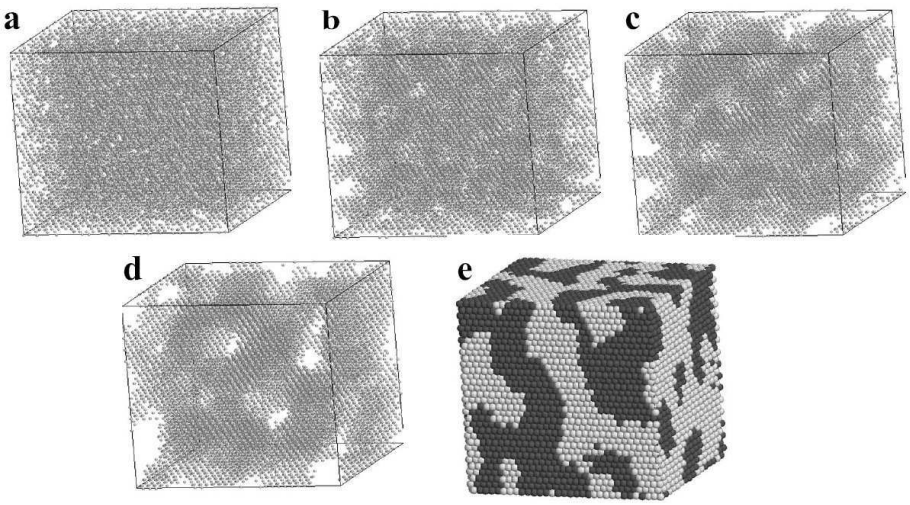

Figure 2: Spinodal decomposition in $Ga_{0.5}In_{0.5}N$ during LKMC annealing run at 473 K. (a)-(d) The spatial distributions of In atoms after (0, 1, 10 and 100)$\times 10^6$ vacancy jumps. (e) The distribution of both Ga (dark) and (In) light atoms in LKMC simulation cell after 9×10^7 vacancy jumps (Publication III).

based on some externally-determined system-specific parameters which, in the end, connect the simulation system to the real one. This means that the input parameters need to be determined in a reliable and accurate way such as performing DFT calculations. During the simulation the defects are allowed to jump into neighboring lattice sites according to the probabilities corresponding to the current defect configuration. By performing a large number of such simulations with similar initial conditions it is then possible to get a qualitative picture of the trends of the evolution of the system. In figure 2(a)-(d) the time evolution for a system consisting of $Ga_{0.5}In_{0.5}N$ is shown after (0, 1, 10 and 100)$\times 10^6$ jumps (Publication III).

3 Computational studies of defects

3.1 Supercell method

A common approach to the calculation of defect properties is to use the supercell approximation. The defect is created in a finite-sized section of the bulk lattice (the supercell) which is then repeated infinitely in different directions using periodic boundary conditions (fig. 3(a)). The supercell method can also be used to study, for example, surfaces (fig. 3(b)) or molecules (fig. 3(c)). The periodicity allows the use of the many techniques developed for periodic systems. For example, the band structure of the (defect-free) bulk is well described unlike in the cluster method. There the cluster is formed by a finite number of host atoms and the surface can affect both the host band structure and the defect states even for relatively large cluster sizes. Another method used in defect calculations is the Green's function embedding technique where the Green's function of the host material is used to calculate the perturbations caused by the creation of the defect. The implementation of, for example, atomic relaxations is difficult in the Green's function method and, thus, this method is rarely used [4, 22].

The unfortunate consequence of the periodicity of the supercell approach is that, instead of one point defect in pure bulk material, there is now an infinite periodic defect array where the defects (or periodic images of the defect) may interact with each other unless the supercell is sufficiently large. These interactions and different ways to deal with them are discussed in sec. 3.2.2. In general, the supercell method is well-suited to the calculation of the formation energies and transition levels of defects using the techniques presented in the next part of this chapter.

3.2 Formation energies and transition levels

The DFT method presented in chapter 2 together with the supercell approximation allows the calculation of the total energy of a system consisting of electrons and ions. From total energies it is possible to estimate the most stable position for a defect in the host lattice or to calculate the migration energy of a vacancy or impurity. It is also possible to obtain the concentration of a defect in thermodynamic equilibrium by using the formation energy of the defect. The formalism by Zhang and Northrup [23] used for

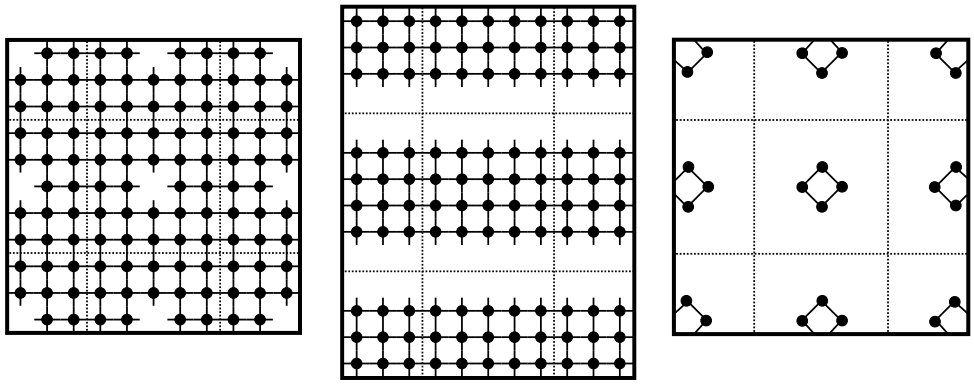

Figure 3: Schematic illustration of supercells and periodic boundary conditions. (a) Vacancy in crystalline solid, (b) slabs for thin film or surface calculations and (c) isolated molecules [4]. Based on [15].

the calculation of the formation energies and transition levels in Publications I, IV and V is presented next following closely the description by Van de Walle and Neugebauer [22].

The concentration of a defect in thermodynamic equilibrium at temperature T is

$$c = N_{\text{sites}} \exp(-E^{\text{f}}/k_{\text{B}}T), \tag{11}$$

where N_{sites} is the number of lattice sites per unit volume where the defect can be formed, E^{f} is the formation energy and k_{B} is the Boltzmann constant. In general, the free energy of formation $F^{\text{f}} = E^{\text{f}} - TS^{\text{f}}$ should be used in eq. (11) instead of the (zero-temperature) formation energy. However, for point defects the formation entropies S^{f} are usually of the order of a few k_{B} and therefore their contributions are not large enough to affect qualitative conclusions.

The formation energy used in eq. (11) can be calculated from the equation

$$E^{\text{f}}[\text{X}^q] = E_{\text{tot}}[\text{X}^q] - E_{\text{tot}}[\text{bulk}] - \sum_i n_i\mu_i + q[E_{\text{F}} + E_{\text{v}} + \Delta V]. \tag{12}$$

where $E_{\text{tot}}[\text{X}^q]$ is the total energy of the supercell where the impurity has been placed into or the defect with charge q created. $E_{\text{tot}}[\text{bulk}]$ is the total

energy of a corresponding supercell containing only bulk material. The chemical potentials μ_i of the n_i atoms added to (n_i positive) or removed from (n_i negative) the supercell are used to take into account the energy of the reservoirs with which the atoms are exchanged. They provide the link to the growth conditions and can be varied within limits which depend on the system in question. They are discussed in more detail in section 3.2.1. The last term of eq. (12) represents the energy of the electron reservoir with which electrons are exchanged. E_{F} is the Fermi energy which is defined here relative to the top of the valence band (E_{v}) of the bulk material. ΔV is a correction term which is used in the alignment of the potentials of the defect and bulk supercells (see sec. 3.2.2).

As can be seen from eq. (12) the formation energy of a charged ($q \neq 0$) defect depends on the Fermi level position in the band gap. The thermodynamic transition level $\epsilon(q_1/q_2)$ (often called the ionization level) is defined as the Fermi level position where the total energies of two charge states q_1 and q_2 become equal. In other words, it is the Fermi energy value where the thermodynamically stable charge state changes. The transition levels are one of the most important properties used in the experimental identification of defects and impurities which almost always induce levels in the band gap. The connection between transition levels and experiments is further discussed in section 3.4.

The formation energies are often shown as a function of the Fermi level while suitable fixed values are chosen for the chemical potentials. An example can be seen in figure 4 where the formation energies of substitutional arsenic defects on the N- ($\mathrm{As_N}$) and Ga-sites ($\mathrm{As_{Ga}}$) in GaN are shown under Ga-rich conditions (Publication I). Each line segment with different slope represents the charge state that has the lowest formation energy for (and thus is the stable charge state at) the Fermi energy value in question. The kinks in the curves correspond to the transition levels.

3.2.1 Chemical potentials

One of the most important questions to consider when calculating formation energies using eq. (12) is how to choose the chemical potentials μ_i for the different elements. The chemical potentials reflect the experimental growth conditions which for the constituent atoms of a III-V compound semiconductor can usually be III- or V-element-rich or somewhere in between. For example, they can be Ga- or N-rich in the case of GaN which is

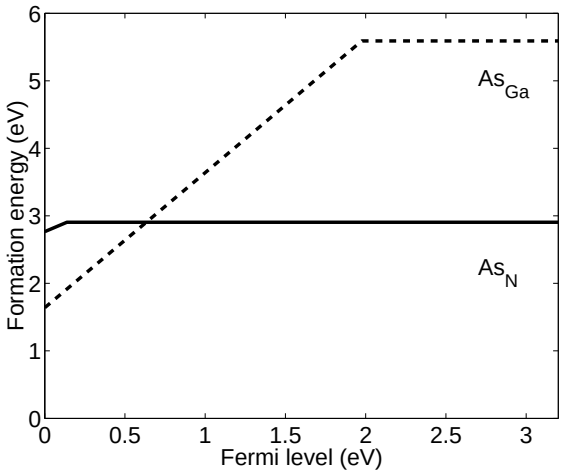

Figure 4: Formation energies of arsenic substitutional defects in wurtzite GaN under Ga-rich conditions (Publication I).

used here as an example of the III-V compound semiconductors discussed in this thesis. The chemical potentials can be considered as variables (with certain bounds) and, unless it is clear that the growth happens always under particular conditions, the results are often reported for both extreme cases. The bounds are usually set by the chemical potentials of the elements in their elemental bulk form (such as bulk Ga and N_2 molecule) together with the chemical potential of the compound semiconductor (in thermodynamic equilibrium):

$$\mu_{\mathrm{GaN}} = \mu_{\mathrm{Ga}} + \mu_{\mathrm{N}}. \tag{13}$$

Under Ga-rich conditions the chemical potential of Ga is equal to the chemical potential of bulk Ga which represents the upper limit after which it becomes energetically more favorable to precipitate pure Ga instead of GaN. The upper bound for N comes from the N_2 molecule in a similar fashion. The lower limits can be obtained by using eq. (13), e.g. the chemical potential of N has its lowest value under Ga-rich conditions and can be calculated from eq. (13) using the chemical potentials of bulk GaN and bulk Ga.

The chemical potentials of the impurity atoms can be more difficult to determine reliably. The chemical potential should reflect the concentration during growth since, for example, if there are no impurity atoms (representing the lower bound of minus infinity), there cannot be any impurity related

defects. An upper bound is given by the elemental bulk phase. Usually there are more stringent bounds due to possible formation of a compound with one of the constituents of the host compound. For example, in the case of N interstitials in GaAs (Publication IV), the N chemical potential can be estimated using either N_2 molecule or bulk GaN.

3.2.2 Finite-size effects

The finite size of the supercell together with the periodic boundary conditions leads to artificial interactions between the periodic images of the defects. For example, the elastic interactions can restrict the relaxation around the defect and the overlap of the defect-state wavefunctions and their periodic images can lead to spurious dispersion of the defect levels.

Point defects can be in different charge states in a semiconductor material and the formation energy of the defect depends on the charge state. Since the periodic images of the defect interact, the total charge of the supercell needs to be zero in order to avoid the divergence of the overall electrostatic energy. The standard way of dealing with this problem is to compensate for the excess charge of the defect by adding a neutralizing uniform background with an opposite charge. The contribution from this artificial charge distribution to the total energy is called the Madelung energy and several different ways to estimate it have been suggested.

If the supercell size could be increased sufficiently, the electrostatic interactions of the system described above would, in principle, become negligible. However, since the interactions converge very slowly (as L^{-1} or L^{-3} where L is the linear dimension of the supercell) with respect to supercell size, the contribution to the total energy can be significant for the currently feasible supercell sizes, depending on the charge distribution in the supercell. One of the most widely known correction schemes is based on the multipole expansion of defect charge distribution where an array of localized charges is immersed in dielectric neutralized by a uniform background charge [24, 25]. However, the derived correction formula, also known as Makov-Payne equation, has been found to be insufficient [26] or to overestimate the correction in some cases [27].

One way to correct for the supercell-size dependent errors from different sources is finite-size scaling. The errors from both elastic and electrostatic interactions of the periodic images of the defects can be expected to have

either L^{-1} (defect-defect image distance) or L^{-3} (number of atoms etc.) dependence so that if the defect calculations were carried out for a number of different sized supercells with the same symmetry, the results could be finite-size scaled and extrapolated to the infinite-sized supercell limit. The scaling method was used, for example, in [26] where it was compared with the potential realignment scheme which is described below and has been used in most of the studies of this thesis. In [26] also the realignment method was found to be successful in giving reliable formation energies even for relatively small supercells. In fact it was found to compensate also for the electrostatic errors even though this is not the usual justification (described below) for its use. In Publication V the dependence of the formation energy on the inverse volume was used to estimate the formation energy in infinite crystal for vacancies in wurtzite AlN and GaN.

Creating a defect in a supercell causes a shift in the potential due to the use of finite-sized supercells and periodic boundary conditions. Therefore an alignment procedure is often used where the potential far from the defect in the defect-containing supercell is aligned with the potential in the bulk supercell. In eq. (12) the potential correction ΔV is included in the term that gives the energy of the electron reservoir with which the electrons are exchanged with i.e. the Fermi energy which is referenced to the valence band maximum E_{v} in the bulk calculation. Simply using the valence band maximum (VBM) of the defect calculation is not possible because the defect strongly affects the band structure.

In Publication IV two different methods of obtaining the potential correction were compared. In the conventional method, which is commonly used in the studies of defects, the correction is obtained as a difference of potential values calculated in an interstitial region as far from the defect as possible (keeping in mind the periodicity) for the defect supercell and for an equivalent interstitial position in the bulk supercell. The interstitial regions are used because of the smoothness of the effective potential which means that the changes in the potential caused by the relaxation of atom positions in the defect supercell are minor compared to using an atomic site where the potential has large variations [28]. The potential values in different supercells are usually estimated by averaging over, for example, $3 \times 3 \times 3$ set of points on a fine grid (the grid was, for example, $168 \times 168 \times 168$ points for a 216-atom supercell in Publication IV). The second method uses the average electrostatic potential at the core of an ion which is calculated by the VASP

code by placing a test charge with norm 1 at each atom and calculating

$$V_{\mathrm{c}}(\mathbf{R}_n) = \int V(\mathbf{r})\rho_{\mathrm{test}}(|\mathbf{r} - \mathbf{R}_n|)\mathrm{d}\mathbf{r}^3 \tag{14}$$

where the radius of the test charge distribution is related to the PAW core radius and kept fixed for each atom type [29]. The shift in the potential can then be calculated by subtracting the potentials of an atom far from the defect in the defect supercell and a corresponding atom in the bulk supercell. After comparison, the latter method of using the average electrostatic potentials at the cores of the ions was found to be as reliable as the commonly used but slightly more laborious method of comparing the potentials of the interstitial regions.

3.3 Defect geometry

A defect changes the local geometry of the host material near its position. This needs to be considered in the calculation of, for example, the formation energies since they would be significantly higher if the defect was not allowed to relax. The relaxation can cause the atoms surrounding the defect to move inwards or outwards and change the symmetry of the defect. In this thesis the geometry of the defect in equilibrium has been determined using the Hellmann-Feynman forces (see section 2.1.4). In practical calculations the Born-Oppenheimer energy surface can be shallow with several competing minima and care is required in the choosing of the computational parameters (including supercell size and shape) and the initial positions of the ions (cf. N interstitials in GaAs, Publication IV). In principle, the defect geometry also offers an opportunity of a straightforward comparison with the experiments since the coordinates of the ions can be directly obtained from the simulation.

In some cases the spatial symmetry of the defect (including the surrounding atoms) is spontaneously lowered by the Jahn-Teller effect which leads to the lowering of the total energy by the removal of electron degeneracy [4]. The negative-U effect happens when the energy gain from the Jahn-Teller distortion exceeds the Coulomb repulsion U between two localized electrons at the defect. For example, for As impurities in GaN (see fig. 4) the charge state changes from +2 directly to 0 with the +1 state not stable for any value of the Fermi energy.

3.4 Comparison with experiments

Theoretical and computational methods, such as the ones described in chapter 2, have become an important tool for the interpretation of experimental results for defects. Many of the electronic and structural properties of defects can be addressed with computational techniques as well as directly or indirectly measured by experiments. Some examples, where the comparison is relatively straightforward, are given in the following.

Calculated formation energies can be used to predict the defects that are more likely to be formed than others under given conditions. They can also be used to predict the charge state and behavior (acceptor or donor) of the defect depending on the doping level in the material.

The thermodynamic transition levels, as defined above, can be directly compared to transitions observed in, for example, deep-level transient spectroscopy (DLTS) or (in some cases) temperature-dependent Hall experiments since in the calculations the final state is allowed to relax to its equilibrium configuration. The optical levels can be defined in a similar way except for the final state which is not allowed to relax but stays in the atomic configuration of the initial state. They can be observed in, for example, photoluminesence experiments. While performing the comparison it is important to be careful and keep in mind the LDA band gap error, which affects the computational results, especially for levels with conduction band character.

Other important quantities that can be compared directly are the hyperfine parameters which can be computationally obtained using wavefunctions from a spinpolarized calculations. They can be compared with results from electron paramagnetic resonance (EPR) and electron-nuclear double-resonance (ENDOR) experiments which give information about the microscopic structure around the defect [30]. Also the calculated frequencies of the localized vibrational modes, which are often induced by defects, can be used in defect identification [31, 4]. The results from various positron annihilation (PA) techniques for open volumes and chemical environments of some defects (in particular, vacancies) can be compared with theoretical predictions.

The densities of states derived from calculations and ballistic electron emission microscopy (BEEM) experiments can be compared as shown, for example, in Publication IV. Multiscale simulation combining DFT and LKMC

in Publication III produced various phase decomposition patterns in GaInN that were compared with transmission electron microscopy (TEM) and X-ray diffraction (XRD) measurements.

3.5 Beyond standard DFT

The standard DFT theory, introduced in the previous chapter, is inherently a ground-state theory and does not allow the calculation of excited-state properties. In particular, the Kohn-Sham wavefunctions and eigenvalues cannot be straightforwardly identified as excited states and should be considered mainly as mathematical tools with a physical meaning only in some rare cases. As described in section 2.1.3, the Kohn-Sham band gap seriously underestimates the fundamental band gap which can cause difficulties especially in the determination of the transition levels. The problem affects both the band structure (the Kohn-Sham eigenvalues) and the total energy calculations since the defect levels in the band gap contribute to the total energy if they are occupied with electrons. The underestimation of the band gap may even lead to some of the defect states that should lie in the gap below the conduction band minimum (CBM) to become resonant with the conduction band instead. Several different methods have been suggested to overcome the problem and some of the most promising ones for defect calculations in semiconductors are discussed next.

One of the simplest schemes is the so-called "scissor operator" [32] where the VBM and CBM from the theoretical calculations are aligned with the VBM and CBM from the experiments stretching the theoretical band gap to the size of the experimental one. In defect calculations this leaves the question of which defect levels should be shifted and how much. Also the self-consistency between the Kohn-Sham eigenvalues and eigenfunctions is lost which may affect the results derived from them [4]. The method has been used in Publication V for neutral and negatively charged N vacancies in GaN and AlN.

Another method based on standard DFT calculations is the so-called "marker" method for calculating transition level positions [33, 34]. The main idea is to estimate the transition level of a defect relative to another experimentally measured or otherwise accurately known transition level of a "marker" defect. The energies of both transfer levels are calculated using the same computational method and parameters which should lead to a

systematic cancellation of computational errors. The reference energy may also be the VBM or CBM of a bulk supercell with the same number of atoms as the defect supercell. The accuracy of the method depends strongly on the choice and availability of an appropriate reference defect [4].

In the "LDA+U" method an additional orbital-dependent interaction is introduced to LDA or GGA calculations. The on-site repulsion is added only to highly localized atomic-like orbitals with the effect of shifting the localized orbitals relative to the other orbitals. This attempts to correct some of the errors of LDA or GGA calculations, such as the band-gap underestimation [35, 36, 37, 4].

The "SIC" (self-interaction correction) approaches attempt to correct the unphysical self-interaction, which is discussed in section 2.1.3 in the context of LDA and GGA exchange-correlation functionals. The corrections are done, for example, by subtracting the self-interaction energy orbital-by-orbital but this can, however, leave some ambiguity as to which states the correction should be applied [38, 4]. An approach based on the self-interaction and relaxation-corrected (SIRC) pseudopotentials offers bulk band structures in good agreement with experiments and improves the defect induced transition levels in the band gap. However, at the moment the method does not allow self-consistent calculation of total energy and thus the effects of the band structure changes on the defect total energies need to be estimated [39, 22, 9].

The Hartree-Fock method, where the exchange is calculated exactly but the correlation is not included, has the opposite problem compared to DFT, namely, it overestimates the band gap. Thus one possible way of improving the band gap in DFT calculations would be to use a so-called hybrid functional where a fraction of exact exchange is mixed with GGA exchange and correlation [5, 40, 41]. The success of the scheme depends on which type of hybrid functional is used. For example, B3LYP [42] appears to predict band gap energies that are within 10% at most of the experimental values for wide range of solids while failing badly on others [10]. The hybrid method is, in general, computationally extremely heavy and is in practise impossible to use for supercells of the sizes required in the defect calculations without further approximations.

Some methods, such as the many-body "GW" quasiparticle theory [43, 44], are mainly aimed at calculating the (quasiparticle) band structure. However, they are prohibitively expensive for large supercells and they are not routinely used in self-consistent total-energy calculations [4].

Quantum Monte Carlo (QMC) methods have only recently been applied to the studies of defect properties (for example, fixed-node diffusion QMC to self-interstitials in Si [45]). They are suitable for total-energy calculations but, for example, the lack of the calculation of forces allowing relaxation and the high computational cost restrict their use in defect calculations for semiconductors [4].

It seems that there is currently no method available that would go beyond standard DFT and give both reliable total energies and be fast enough for defect calculations. However, there are many examples of state-of-the-art calculations based on the standard DFT methods which provide reliable information on the properties of defects in semiconductors and thus contribute to their identification and to the knowledge of the processes that control them (for example, see [22]).

4 Results

4.1 GaAsN alloys

One of the most attractive properties of GaAsN alloys is that they can, in principle, be used to produce light of wavelength ranging from infrared (GaAs) to ultraviolet (GaN) because of the large difference of the band gap energies of GaAs and GaN. However, only As-rich (low concentration of N in GaAs) or N-rich (low concentration of As in GaN) alloys can be reached experimentally because of the large miscibility gap [46, 47] which is caused by the large lattice constant mismatch between GaAs and GaN [48]. The large mismatch in size and electronegativity between nitrogen and arsenic atoms also causes other unusual properties compared to more conventional III-V semiconductor alloys [49]. One of the most important effects is the dramatic decrease of the band gap with the introduction of just a few percent of nitrogen in GaAs [50, 49]. The reduction slows down for higher N concentrations which means that the band gap bowing parameter has a strong composition dependence for small N concentrations [47, 51]. For the GaAsN alloys the bowing effect is very strong which allows large changes in the band gap value with a small change in composition [50]. As-rich $GaAs_{1-x}N_x$ alloys have already been studied extensively both experimentally and theoretically but there are only few theoretical studies on the N-rich $GaN_{1-y}As_y$ alloys.

Usually it is assumed that in alloys elements will mostly substitute atoms of the same group (for example, N replaces As since both are group-V elements) i.e. go to isoelectric substitutional sites. Although this also applies to As-rich GaAsN where most of the nitrogen atoms occupy As-sites, in several experimental studies it has been shown that single N atoms and N-N dimers can also be found in different interstitial sites [52, 53, 54]. In Publication I the isoelectric substitutional sites have been assumed while the interstitial N has been studied in Publication IV. In Publication I also the occupation site of As (substitutional N- or Ga-site) in GaN was investigated.

4.1.1 As- and N-rich GaAsN alloys

In Publication I we studied the atomic and electronic structures of As-rich $GaAs_{1-x}N_x$ and N-rich $GaN_{1-y}As_y$ alloys in a large composition range. We

investigated systematically the effect of the impurity atom configuration near both GaAs and GaN sides of the concentration range on the total energies, lattice constants and band gaps.

Firstly, the alloying has a strong effect on the local atomic structures of GaAs and GaN because of the size difference of the isoelectric N and As atoms. The N atoms, when they substitute As atoms in GaAs, cause the surrounding Ga atoms to relax inwards, making the Ga-N bond length about 15% shorter than the corresponding Ga-As bond length in GaAs. In contrast to this, the As atoms substituting N atoms in GaN cause the surrounding Ga atoms to move outwards, so that the Ga-As bond length becomes about 15% longer than the corresponding Ga-N bond length in GaN. Secondly, the different configurations of the alloying atoms in the supercells cause the total energies of the relaxed alloy supercells and the band gaps to experience large fluctuations which grow stronger if the impurity concentration is increased. Substituting As atoms with N in GaAs induces modifications near the conduction band minimum while substituting N atoms with As in GaN modifies the states near the valence band maximum. Both lead to band gap reduction which is first rapid but later slows down. The relative size of the fluctuations is much larger in the case of $GaAs_{1-x}N_x$ alloys.

We also looked into the question of which site, substitutional N or Ga, As would prefer in GaN and how to interpret some of the experimental findings. We calculated the formation energies and the transition levels of the two possible arsenic substitutional defects in GaN and found that the As atoms prefer the N substitutional site over the Ga site under Ga-rich conditions within a large range of Fermi level values. The As_N defect has a transition level at 0.14 eV above the valence band maximum in wurtzite GaN and is the most likely cause for the near band-edge emission found in many experiments.

4.1.2 N interstitials in GaAs

As discussed above, the nitrogen atoms in dilute nitrogen GaAsN can occupy interstitial sites as well as substitutional sites as has been shown previously by many experiments. In Publication IV we calculated the formation energies and transition levels of the most important interstitial defect types in different charge states using the method described in the previous chapters. As can be seen from fig. 5, the lowest-energy interstitial-type defects were

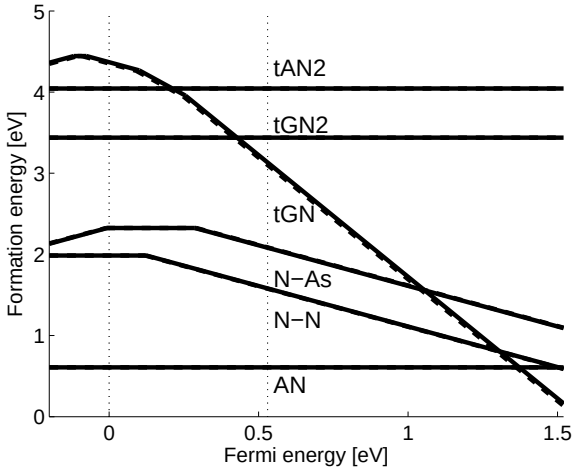

Figure 5: The formation energies of various N interstitial defects and N on substitutional As-site (AN) in GaAs as a function of Fermi energy under As-rich conditions. In the picture the comparison of the two different methods of calculating the potential correction is also shown. The results obtained using the interstitial regions are shown with solid lines and the core regions with dashed lines. Only those line segments corresponding to the charge state with the lowest energy at each Fermi energy value are shown with the kinks corresponding to the transition levels. For the notation and structures of the different types of interstitial defects see fig. 6.

found to be N-As (fig. 6(a)) and N-N (fig. 6(b)) split interstitials for most of the experimentally relevant conditions. Their most important transition levels in the band gap are $(0/-)$ transitions at 0.12 eV and 0.29 eV above the valence band maximum for the N-N and N-As defects, respectively.

We also performed a comparison of two different methods for determining the potential correction used in the calculation of the formation energies and found that the difference between the results given by the new method of using the average electrostatic potentials of the cores of the ions and the conventional method of using the interstitial regions is small enough to be insignificant both qualitatively and quantitatively (see fig. 5 and section 3.2.2).

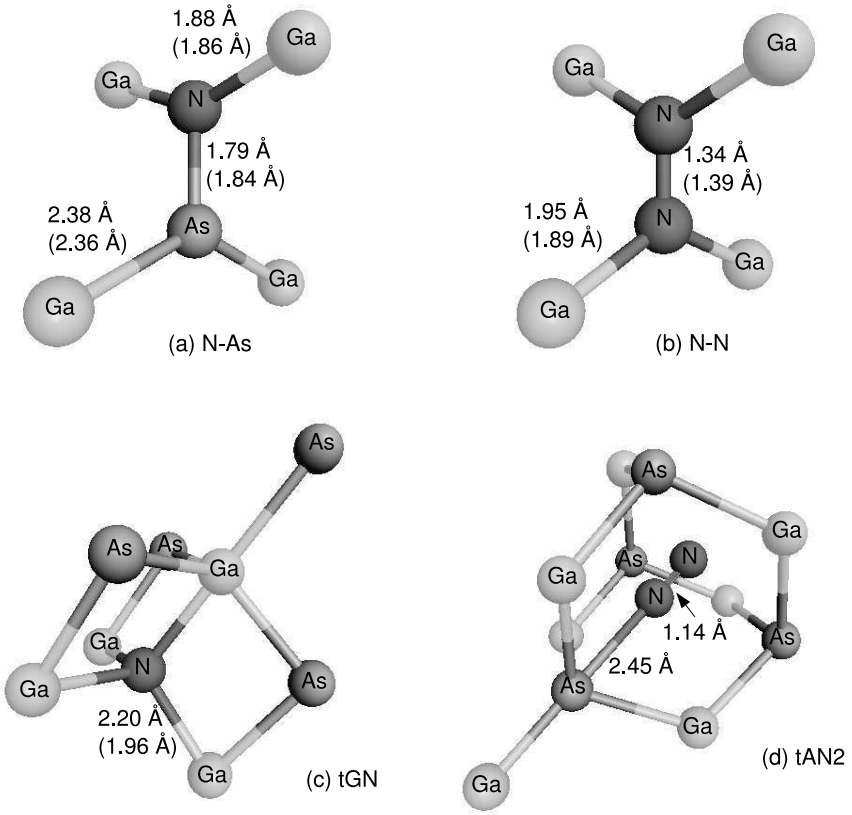

Figure 6: The structure of (a) N-As and (b) N-N split interstitial defects and (c) tGN and (d) tAN2 (in tGN2 As and Ga atoms exchange positions) defects in the neutral charge state. Bond lengths shown in brackets correspond to the dominant charge state which is -1 for both N-As and N-N and -3 for tGN.

4.2 GaInN alloys

The III-V semiconductors with N as the fifth-group element have unique properties that make them very attractive for short-wavelength light emitters and high-power/high-temperature electronics applications [55]. GaInN alloys have an ability to emit light in the wavelength range from ultraviolet (GaN) to red (InN). However, in experiments they have been shown to have a tendency to undergo phase decomposition which can cause problems in applications since the non-uniformity of composition of the alloy can have an unfavorable effect on its electronic properties [56].

There are indications that the thermodynamic instability is an inherent property of wurtzite GaInN (instead of a growth artifact) but the formation of the microstructure in GaInN is not yet fully understood. GaInN demonstrates multiple, sometimes seemingly mutually contradictory, decomposition modes, including alloy ordering [57, 58] and phase separation: either precipitation [59, 60], or compositional modulations [61, 62] which have even been observed simultaneously in the same alloy samples in some experiments [63].

The reasons behind the decomposition have been extensively investigated before in many theoretical studies (see, for example, [64, 65, 66] and references therein). However, many of them have not taken into account the difference between the zinc-blende and the (ground-state) wurtzite modifications of GaInN (in wurtzite the atoms have more non-equivalent neighbors, see fig. 7) or the local and external elastic strain (former caused by the difference in size between In and Ga atoms and the latter, for example, by growth on a substrate with different lattice parameters). It is also important to obtain adequate estimates for the material-dependent parameters (such as interaction energies) from, for example, first-principles calculations.

In Publications II and III we investigated the minor component ordering in wurtzite $Ga_{1-x}In_xN$ ($x \leq 0.5$) alloys. We performed total-energy density-functional calculations, which were used to investigate the interaction of Ga and In atoms in GaInN and its sensitivity to variations of alloy composition x and to external strains. In Publication III the DFT results were used to simulate the decomposition of an initially random alloy using lattice kinetic Monte Carlo. With the help of such multiscale modeling we demonstrated the decomposition patterns corresponding to the ab-initio values of metal atom interaction parameters and predicted also how these patterns would be modified, provided these parameters were varied within reasonable limits.

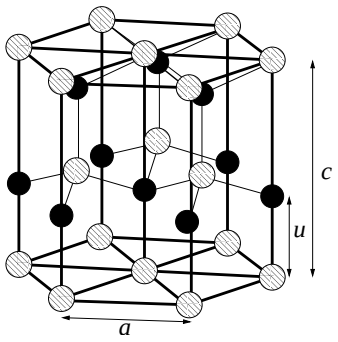

Figure 7: Schematic representation of wurtzite structure. Black circles represent N atoms and shaded circles represent Ga or In atoms.

The simulations demonstrated that the phase decomposition patterns in wurtzite GaInN are very sensitive to the interplay of metal atom interactions at several interatomic distances at least to the fourth nearest neighbors on the cation sublattice. Several different approaches were used to estimate the interaction energies based on DFT calculations and they all agree that the differences between these energies are quite small. However, the absolute values and even the signs of the binding energies of In atoms were found to depend on the assumed effects of the local chemical composition and the external elastic strains. The variation of the metal interaction energies within reasonable limits (based on the estimations above) and the concentration of In in the alloy resulted in very different relaxation patterns such as linear or wall ordering of In and Ga atoms along *c*-axis, planar ordering parallel to basal plane and spinodal decomposition (see figure 2). This high sensitivity of the GaInN decomposition to relatively small variations of the metal interaction energies could be the main reason for the experimentally observed versatility of the alloy decomposition patterns and their sensitivity to the particular experimental conditions.

In addition to the GaInN alloys we also investigated GaAlN alloys in Publication II in order to compare our results for GaInN with another III-nitride alloy. In contrast to In in GaN, Al atoms in GaN do not interact irrespective of strain conditions and can be expected to form a random alloy. In the same paper, the effect of the minor component (In or Al) ordering pattern on the band gap of GaInN and GaAlN alloys was also considered and the band gap was found to be dependent on the In atom configuration in GaInN.

4.3 Vacancies in AlN and GaN

Both GaN and AlN are wide-band-gap semiconductors with promising applications in the blue-ultraviolet optoelectronics. Among all III-nitrides, wurtzite AlN has the widest direct gap ($E_g = 6.1$ eV) which makes it very attractive when attempts are made for moving the optical emission and/or detection wavelength deeper into the UV region. One of the problems encountered in the development of devices is n- and p-type doping which has proven to be challenging in both GaN and AlN.

Since defects can significantly affect the material properties, the native defects in these materials have been for long intensely studied. In particular, N vacancy in GaN was first revealed by calculations to have too high formation energy in n-type material for them to exist in abundance in thermal equilibrium [67, 68, 69]. Ga vacancies (V_{Ga}) were considered to be the only native defects to exist in observable amounts in as-grown GaN. However, recently it was suggested that nitrogen vacancies, including negatively charged ones, can be major participants in the defect kinetics also in n-type GaN [70].

Most of the computational studies of defects in GaN and AlN, to our knowledge, have been performed for the zinc-blende structure, even though the main polytype of both materials is wurtzite. In AlN there is an important difference in the behavior of the defect levels induced by N vacancy in zinc-blende and wurtzite AlN. In zinc-blende the upper defect levels are in resonance with the conduction band while in wurtzite they lie well below the conduction band edge [71]. The negative charge states of the N vacancy are not considered in many studies (for example, in [72]) while other studies have given conflicting results concerning their importance for the material properties [73, 74].

The calculations are usually performed using periodically repeated finite-sized supercells which introduces artificial long-range elastic and electrostatic interactions between the periodic defect images. The finite-size effects are stronger for small supercells and many of the previous calculations have been performed for supercells with less than 96 atoms. In Publication V we investigated the dependence of the formation energy on the supercell size for charged vacancies in GaN and AlN. The dependencies were used to estimate the formation energies of the vacancies in infinite crystal.

Since GaN and AlN are wide-band-gap materials, it is important to take into account the effects of the DFT underestimation of the band gap. Before the calculation of the formation energies using eq. (12), we used the

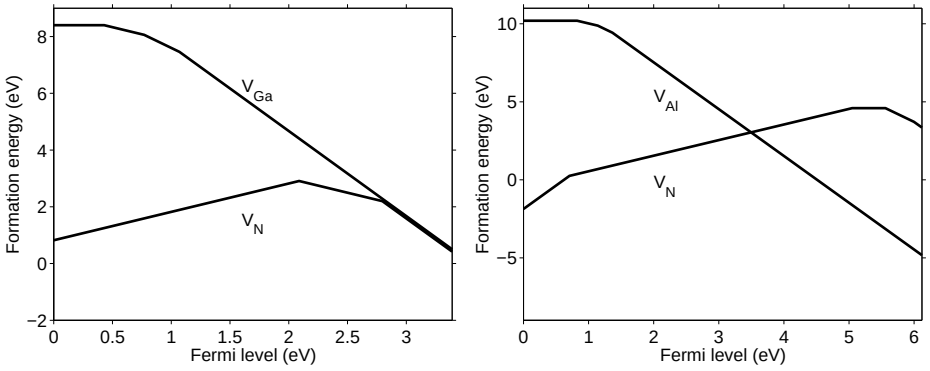

Figure 8: The formation energies of vacancies in GaN (left) and AlN (right) as a function of the Fermi level under Ga- and Al-rich conditions, respectively.

so-called "scissor"-operator (see section 3.5 and Publication V for details) to correct the total energies of the defect supercells affected by the band-gap underestimation. In our case the scissor operator was applied to the occupied defect levels near CBM for the neutral and negatively charged N vacancies but not to the defect states derived from the valence band states.

The formation energies of vacancies in GaN and AlN are shown in figure 8. The results for the cation vacancies agree well with previous results. The N vacancies in GaN are found to be important defects at all Fermi energy values. In particular, in -3 charge state they have similar formation energies with Ga vacancies in n-type GaN. In AlN the N vacancy has stable negative charge states and compared to the results from the previous studies [73, 74] the transition levels involving them are located quite near to the conduction band minimum. However, the formation energy of the N vacancy is rather high in n-type material which implies that their concentration is low in thermal equilibrium.

For the N vacancies we also considered the localization of the electrons on the highest occupied states since some of the corresponding total-energy derived transition levels lie above the computational conduction band minimum. They were found to be localized in the defect region in both GaN and AlN which suggests that the levels are located in the upper part of the band gap in both materials.

5 Summary

This thesis presents computational studies of the III-V compound semiconductors which are essential in many electronics and optoelectronics applications such as white light emitting diodes (LEDs) and blue lasers. The electronic and structural properties of the materials are greatly affected by defects in binary compounds and composition uniformity in ternary alloys.

The results include the determination of the prevalence and properties of different types of defects in GaN, AlN and GaAs compound semiconductors. In GaN, the As defects are found to substitute N atoms and to be the likely source of the near-band-edge emission found in many experiments. The most common N interstitial defects in GaAs are shown to be the N-N and N-As split interstitials. The nitrogen vacancies in wurtzite GaN are found to be more important to the defect kinetics than what has been previously assumed.

The alloy studies clarify the effects of alloying in As- and N-rich GaAsN alloys. The multiscale modeling of GaInN reveals the high sensitivity of the GaInN decomposition to relatively small variations of the interaction between Ga and In atoms offering an explanation for the diversity of the alloy decomposition patterns observed in the experiments.

The DFT methods used in the majority of the studies included in this thesis are well-suited for studying the effects of the defects on the properties of the III-V compound semiconductors. Some of their existing limitations might be overcome by the development of the numerical algorithms and computers while others by the development of methods which, while at the moment computationally too heavy for practical calculations for defects, show great promise for the future.

References

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[3] A. E. Mattsson, Science **298**, 759 (2002).

[4] R. M. Nieminen, in *Theory of Defects in Semiconductors*, edited by D. A. Drabold and S. Estreicher (Springer, Berlin/Heidelberg, 2006).

[5] J. P. Perdew and S. Kurth, in *A Primer in Density Functional Theory*, edited by C. Fiolhais, F. Nogueira, and M. Marques (Springer, Berlin, 2003).

[6] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).

[7] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[8] J. P. Perdew, M. Ernzerhof, A. Zupan, and K. Burke, J. Chem. Phys. **108**, 1522 (1998).

[9] C. Stampfl, C. G. Van de Walle, D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **61**, R7846 (2000).

[10] S. Kümmel and L. Kronik, Rev. Mod. Phys. **80**, 3 (2008).

[11] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982).

[12] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[13] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[14] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[15] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).

[16] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[17] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[18] H. Hellmann, *Einführung in die Quantenchemie* (Deuticke, Leipzig, 1937).

[19] R. P. Feynman, Phys. Rev. **56**, 340 (1939).

[20] M. G. Ganchenkova, V. A. Borodin, and R. M. Nieminen, Nucl. Instrum. Methods Phys. Res. B **228**, 218 (2005).

[21] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999).

[22] C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).

[23] S. B. Zhang and J. E. Northrup, Phys. Rev. Lett. **67**, 2339 (1991).

[24] M. Leslie and M. J. Gillan, J. Phys. C **18**, 973 (1985).

[25] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).

[26] C. W. M. Castleton, A. Höglund, and S. Mirbt, Phys. Rev. B **73**, 035215 (2006).

[27] J. Shim, E.-K. Lee, Y. J. Lee, and R. M. Nieminen, Phys. Rev. B **71**, 035206 (2005).

[28] T. Mattila and A. Zunger, Phys. Rev. B **58**, 1367 (1998).

[29] H.-P. Komsa, E. Arola, E. Larkins, and T. T. Rantala, J. Phys.: Condens. Matter. **20**, 315004 (2008).

[30] C. G. Van de Walle and P. E. Blöchl, Phys. Rev. B **47**, 4244 (1993).

[31] C. G. Van de Walle, Phys. Rev. Lett. **80**, 2177 (1998).

[32] G. A. Baraff, E. O. Kane, and M. Schlüter, Phys. Rev. Lett. **43**, 956 (1979).

[33] J. Coutinho, V. J. B. Torres, R. Jones, and P. R. Briddon, Phys. Rev. B **67**, 035205 (2003).

[34] J. Coutinho, S. Öberg, V. J. B. Torres, M. Barroso, R. Jones, and P. R. Briddon, Phys. Rev. B **73**, 235213 (2006).

[35] V. I. Anisimov, F. Aryasetiawan, and A. I. Liechtenstein, J. Phys.: Condens. Matter **9**, 767 (1997).

[36] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2004).

[37] A. Janotti and C. G. Van de Walle, Phys. Rev. B **76**, 165202 (2007).

[38] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[39] D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **55**, 12836 (1997).

[40] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[41] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[42] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).

[43] S. Ismail-Beigi and S. G. Louie, Phys. Rev. Lett. **95**, 156401 (2005).

[44] M. Hedström, A. Schindlmayr, G. Schwarz, and M. Scheffler, Phys. Rev. Lett. **97**, 226401 (2006).

[45] W.-K. Leung, R. J. Needs, G. Rajagopal, S. Itoh, and S. Ihara, Phys. Rev. Lett. **83**, 2351 (1999).

[46] C. T. Foxon *et al.*, J. Cryst. Growth **150**, 892 (1995).

[47] W. G. Bi and C. W. Tu, Appl. Phys. Lett. **70**, 1608 (1997).

[48] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **51**, 10568 (1995).

[49] L. Bellaiche, S.-H. Wei, and A. Zunger, Appl. Phys. Lett. **70**, 3558 (1997).

[50] I. Vurgaftman and J. R. Meyer, J. Appl. Phys. **94**, 3675 (2003).

[51] J. Salzman and H. Temkin, Mater. Sci. Eng. B **50**, 148 (1997).

[52] W. Li, M. Pessa, and J. Likonen, Appl. Phys. Lett. **78**, 2864 (2001).

[53] T. Ahlgren, E. Vainonen-Ahlgren, J. Likonen, W. Li, and M. Pessa, Appl. Phys. Lett. **80**, 2314 (2002).

[54] W. J. Fan, S. F. Yoon, T. K. Ng, S. Z. Wang, W. K. Loke, R. Liu, and A. Wee, Appl. Phys. Lett. **80**, 4136 (2002).

[55] S. J. Pearton, J. C. Zolper, R. J. Shul, and F. Ren, J. Appl. Phys. **86**, 1 (1999).

[56] N. Wieser, O. Ambacher, H.-P. Felsl, L. Görgens, and M. Stutzmann, Appl. Phys. Lett. **74**, 3981 (1999).

[57] P. Ruterana, G. Nouet, W. Van der Stricht, I. Moerman, and L. Considine, Appl. Phys. Lett. **72**, 1742 (1998).

[58] D. Doppalapudi, S. N. Basu, K. F. Ludwig Jr., and T. D. Moustakas, J. Appl. Phys. **84**, 1389 (1998).

[59] N. A. El-Masry, E. L. Piner, S. X. Liu, and S. M. Bedair, Appl. Phys. Lett. **72**, 40 (1998).

[60] M. D. McCluskey, L. T. Romano, B. S. Krusor, D. P. Bour, N. M. Johnson, and S. Brennan, Appl. Phys. Lett. **72**, 1730 (1998).

[61] Z. Liliental-Weber, D. N. Zakharov, K. M. Yu, J. W. Ager III, W. Walukiewicz, E. E. Haller, H. Lu, and W. J. Schaff, Physica B **376-377**, 468 (2006).

[62] E. Zielinska-Rohozinska, J. Gronkowski, M. Regulska, M. Majer, and K. Pakula, Cryst. Res. Technol. **36**, 903 (2001).

[63] M. K. Behbehani, E. L. Piner, S. X. Liu, N. A. El-Masry, and S. M. Bedair, Appl. Phys. Lett. **75**, 2202 (1999).

[64] E. J. Miller and E. T. Yu, Appl. Phys. Lett. **78**, 2303 (2001).

[65] M. Shimotomai and A. Yoshikawa, Appl. Phys. Lett. **73**, 3256 (1998).

[66] L. K. Teles, J. Furthmüller, L. M. R. Scolfaro, J. R. Leite, and F. Bechstedt, Phys. Rev. B **62**, 2475 (2000).

[67] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50**, 8067 (1994).

[68] T. Mattila, A. P. Seitsonen, and R. M. Nieminen, Phys. Rev. B **54**, 1474 (1996).

[69] T. Mattila and R. M. Nieminen, Phys. Rev. B **55**, 9571 (1997).

[70] M. G. Ganchenkova and R. M. Nieminen, Phys. Rev. Lett. **96**, 196402 (2006).

[71] C. Stampfl and C. G. Van de Walle, Phys. Rev. B **65**, 155212 (2002).

[72] Y. Zhang, W. Liu, and H. Niu, Phys. Rev. B **77**, 035201 (2008).

[73] H.-G. Ye, G.-D. Chen, Y.-Z. Zhu, and H.-M. Lü, Chinese Phys. **16**, 3803 (2007).

[74] A. Fara, F. Bernardini, and V. Fiorentini, J. Appl. Phys. **85**, 2001 (1999).